GREQAM
Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes études en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2011-19

Inequality decomposition using the Gibbs output of a Mixture of lognormal distributions

Michel Lubrano
Abdoul Aziz Junior Ndoye

April 2011

DT-GREQAM

Inequality decomposition using the Gibbs output of a Mixture of lognormal distributions

Michel Lubrano* Abdoul Aziz Junior Ndoye†

April, 2011

Abstract

In this paper we model the income distribution using a Bayesian approach and a mixture of lognormal densities. The size of the mixture is determined by Chib (1995)'s method. Using the Federal Expenditure Survey data for the United Kingdom, we detect three groups corresponding to the three classes (poor, middle class and rich).

The marked growth in UK income inequality during the late 1970s is increasingly attracting attention. The increasing gap between the poorest and the richest was accompanied by changes in the clustering of incomes in between. Using the decomposable Generalised Entropy (GE) inequality indices, we carry out a within-between group analysis of income inequality in the three identified groups in UK during 1979 to 1996 and show the evolution of the importance of each group. Whereas during the late 1970s the concentration of people around middle income levels began to break up and polarise towards high and low incomes as shown by Jenkins (1996), our Bayesian results show that the inequality within the low and middle income group do not change much and the importance of the high income is the most affected by the fight against inequality that followed the Thatcher period.

keywords: Income distribution, Generalised Entropy, Mixture models, Gibbs sampler, Marginal likelihood.

*CNRS, GREQAM Marseille email: michel.lubrano@univmed.fr

†GREQAM, Marseille email: abdoul-aziz-junior.NDOYE@univmed.fr

1 Introduction

Income distributions have been estimated using non-parametric and parametric approaches. Each approach has its own drawbacks. Non parametric methods are fully able to reproduce the details of most empirical samples (provided some caution as in Marron and Schmitz 1992). Kernel density estimation provides a useful pictorial representation of the structure of the data. Nonetheless, they require a delicate choice of the value for the smoothing parameter. Estimated parametric densities have nice statistical properties. New families of densities were explored so as to produce better flexibility in the tails, a point where non parametric methods are weak. Nonetheless, they make an underlying restrictive hypothesis that the distribution is unimodal, so they cannot detect heterogeneity in the sample and inferences about inequality can also depend critically on which distribution is chosen (Singh and Maddala 1976, McDonald and Ransom 1979, McDonald 1984,...). To overcome these restrictions, a functional form that is relatively flexible is needed. Thanks to their semi-parametric framework, mixture models can provide flexible specifications and, under certain conditions, can approximate consistently any form of distributions (Ghosal and Van der Vaart 2001). However, mixtures are sensitive to the choice of the basic density and to the choice of the estimation method (Flachaire and Nunez 2002 and Chotikapanich and Griffiths 2008). Using the Federal Expenditure Survey (FES) data for the United Kingdom, Flachaire and Nunez (2002) have modeled the UK's income distribution using a mixture of log-transformed normal densities estimated by Maximum Likelihood using the EM algorithm. They have identified five groups for each of the four years that cover the FES data set. However, it is difficult to interpret clearly these estimated groups in term of specific social groups or classes.

In this paper, we first propose to model the income distribution by using a finite mixture of lognormal densities directed adjusted on the raw (untransformed) data using a Bayesian approach based on the Gibbs sampler. We solve the switching label problem by eliciting a semi informative prior. We adapt the Chib (1995)'s method to estimate the marginal likelihood in order to select the optimal number of components. Using the same data set as Flachaire and Nunez (2002), we detect three groups for each year corresponding to three social classes: poor, middle class and rich.

Since the late 1970s, real income increased a lot in the UK, but the gap between the poorest and the richest also increased faster than in any comparable industrial countries. During the late 1970s, Jenkins (1996) showed that the distinct clump in the concentration of people around middle income levels began to break up and polarise towards high and low incomes. Mr Kinnock claimed that "While the very rich have lost some of their riches to the less rich, over time, the poor have hardly profited proportionately".[1] By contrast, Mrs Thatcher view was that "the real incomes have increased throughout all income groups". In this paper, we use a class of decomposable inequality indices to illustrate our combination approach by carrying out a within-between group analysis of income inequality in the three groups identified in UK during the period 1979-1996. We consider the popular and leading class of inequality indices, the Generalised Entropy (GE) indices. Our Bayesian results confirm the large change in structure of the three groups. However, we find that the importance of the low income group and the middle income group do not change much, that of the high income is the most affected by the fight against inequality that followed the Thatcher period. We also find that within inequality represents on average 68% of total inequality and that this proportion does not vary much.

The paper is organised as follows. Section 2 provides a Bayesian inference for mixture of lognormal distributions, implements the Chib (1995)'s method and fits income distributions for FES data. Section 3 reviews the GE family index, its decomposition and using the Gibbs output obtained, it gives also Bayesian inference for the GE indices. Section 4 illustrates of the method used. Finally section 5 concludes.

2 Estimation of income distribution using a mixture of lognormal distributions

Aitchison and Brown (1957) argued that the lognormal distribution is particularly convenient for the distribution of incomes in fairly homogeneous subpopulation of the workforce. However, the observed population results from the mixing of various sub-populations. An accurate modeling for this heterogeneity requires the use of mixture models which are shown to be able to consistently estimate any probability density function under some regularity conditions.

2.1 Bayesian inference for mixture of lognormal distributions

A finite mixture of lognormal distributions is a convex combination of k lognormal distributions where the density of the j^{th} component is given by $\Lambda(x|\mu_j, \sigma_j^2)$ with (μ_j, σ_j^2) being the component specific mean and variance. If each component is sampled with probability p_j ($\sum p_j = 1$), then the density function of the data x

[1] The Future of Socialism, Fabian Tract No.506, January 1986.

given the parameters (μ, σ^2, p) is

$$f(x|\mu, \sigma^2, p) = \sum_{j=1}^{k} p_j \Lambda(x|\mu_j, \sigma_j^2), \tag{1}$$

where $\mu = (\mu_1, ..., \mu_k)$, $\sigma^2 = (\sigma_1^2, ..., \sigma_k^2)$, $p = (p_1, ..., p_k)$ and

$$\Lambda(x|\mu_j, \sigma_j) = \frac{1}{x\sigma_j\sqrt{2\pi}} \exp -\frac{(\log x - \mu_j)^2}{2\sigma_j^2}.$$

A sample $(x_1, \ldots, x_n)$ from (1) can be seen as a collection of sub-samples originating from each of the $\Lambda(x|\mu_j, \sigma_j^2)$'s, when both the size and the origin of each sub-sample are unknown. That is, each x_i's in the sample is a priori distributed from any of the Λ_j's with probabilities p_j. Depending on this setting, the inferential goal behind this modeling may be to reconstitute the original homogeneous sub-samples by estimating the number of components k and providing estimators for the parameters μ_j, σ_j^2 and p_j of the different components.

To do this the missing data representation is used to facilitate numerical estimation. For each observation x_i we associate a missing variable z_i that indicates its component. Formally, this means that we have a hierarchical structure associated with the model:

$$z_i|p \sim \mathrm{M}_k(p_1, \ldots, p_k), \qquad x_i|z_i, \mu, \sigma^2 \sim \Lambda(\cdot|\mu_{z_i}, \sigma_{z_i}^2),$$

where $z_i \in \{1, \ldots, k\}$.

For a given allocation size vector $(n_1, \ldots, n_k)$, where $n_1 + \ldots + n_k = n$, we define the partition sets

$$Z_j = \left\{ z : \sum_{i=1}^{n} \mathbf{1}_{z_i=1} = n_1, \ldots, \sum_{i=1}^{n} \mathbf{1}_{z_i=k} = n_k \right\},$$

which consist of all allocations with the given allocation size $(n_1 + ... + n_k)$.

Since the lognormal distribution belongs to the exponential family, we can derive easily some sufficient statistics which make easy the computation of the likelihood function and also make straightforward simulations of parameters.

The conditional likelihood function of a sample is

$$\begin{aligned}
L(\mu_j, \sigma_j^2|x,z) &= \left(\prod_{i\in Z_j}(x_i)^{-1}\right)(2\pi)^{-n_j/2}\sigma_j^{-n_j}\exp -\frac{1}{2\sigma_j^2}\sum_{i\in Z_j}(\log x_i - \mu_j)^2 \\
&\propto \sigma_j^{-n_j}\exp -\frac{1}{2\sigma_j^2}\sum_{i\in Z_j}(\log x_i - \mu_j)^2 \\
&\propto \sigma_j^{-n_j}\exp -\frac{1}{2\sigma_j^2}\left(s_j^2(z) + n_j(\mu_j - \bar{x}_j(z))^2\right), \qquad (2)
\end{aligned}$$

where

$$\bar{x}_j(z) = \frac{1}{n_j}\sum_{i\in Z_j}\log x_i \qquad s_j^2(z) = \frac{1}{n_j}\sum_{i\in Z_j}(\log x_i - \bar{x})^2.$$

In order to carry out a full Bayesian analysis for the mixture of lognormal process, conjugate prior densities of all parameters in the model must be specified. We can thus select a conditional normal prior on μ_j, an inverted gamma2 prior on σ_j^2 and a Dirichlet prior on p

$$\begin{aligned}
\pi(\mu_j|\sigma_j^2) &= f_N(\mu_j|\mu_0, \sigma_j^2/n_0) \propto \sigma_j^{-1}\exp -\frac{n_0}{2\sigma_j^2}(\mu_j - \mu_0)^2, \qquad (3) \\
\pi(\sigma_j^2) &= f_{i\gamma}(\sigma_j^2|\nu_0, s_0) \propto \sigma_j^{-(\nu_0+2)}\exp -\frac{s_0}{2\sigma_j^2}, \qquad (4) \\
\pi(p) &= f_{Dir}(\gamma_1, ..., \gamma_k) \propto \prod_{i=1}^{k} p_i^{\gamma_i - 1}. \qquad (5)
\end{aligned}$$

Let us now combine the prior with the likelihood function to obtain the joint posterior probability density function of $(\mu_j,\ \sigma_j^2)$ in such a way that isolates the conditional posterior densities of each parameter.

$$\pi(\mu_j, \sigma_j^2|x,z) \propto \sigma_j^{-(n_j+\nu_0+3)}\exp -\frac{1}{2\sigma_j^2}\left(s_0 + s_j^2(z) + n\left(\mu - \bar{x}_j(z)\right)^2 + n_0(\mu - \mu_0)^2\right).$$

As we are in the natural conjugate framework, we must identify the parameters of the product of an inverted gamma2 in σ_j^2 by a conditional normal density in $\mu_j|\sigma_j^2$. After some algebraic manipulations: the conditional normal posterior is

$$\begin{aligned}
\pi(\mu_j|\sigma_j^2, x, z) &\propto \sigma_j^{-1}\exp -\frac{1}{2\sigma_j^2}\left((n_0\mu_0 + n_j\bar{x}_j(z))/n_{*j}\right), \\
&\propto f_N(\mu_j|\mu_{*j}, \sigma_j^2/n_{*j}),
\end{aligned}$$

with

$$n_{*j} = n_0 + n_j, \qquad \mu_{*j} = (n_0\mu_0 + n\bar{x}_j(z))/n_{*j}.$$

Then the marginal posterior density of μ_j is Student with

$$\pi(\mu_j|x,z) = f_t(\mu_j|\mu_{*j}, s_{*j}, n_{*j}, \nu_{*j}), \tag{6}$$

where

$$\nu_{*j} = \nu_0 + n_j, \qquad s_{*j} = s_0 + s_j^2(z) + \frac{n_0 n_j}{n_0 + n_j}(\mu_0 - \bar{x}_j(z))^2.$$

The posterior density of σ^2 is given by

$$\begin{aligned} \pi(\sigma_j^2|x,z) &\propto \sigma_j^{-(n_j+\nu_0+2)} \exp -\frac{1}{2\sigma_j^2}\left(s_0 + s_j^2(z) + \frac{n_0 n_j}{n_0 + n_j}\left(\mu_0 - \bar{x}_j(z)\right)^2\right), \\ &\propto f_{i\gamma}(\sigma_j^2|\nu_{*j}, s_{*j}). \end{aligned} \tag{7}$$

Of course, all the posterior densities are conditional on the value of z. This motivates a Gibbs algorithm which involves the successive simulation of z, p, μ and σ^2.

1. Initialize $p^{(0)}$, $\mu^{(0)}$, $\sigma^{2^{(0)}}$, $t = 0$
2. $t = t + 1$
3. Generate $z_i^{(t)}$ $(i = 1, \ldots, n,\ j = 1, \ldots, k)$ from
$$\mathbb{P}\left(z_i^{(t)} = j|p_j^{(t-1)}, \mu_j^{(t-1)}, \sigma_j^{2(t-1)}, x_i\right) \propto p_j^{(t-1)}\Lambda\left(x_i|\mu_j^{(t-1)}, \sigma_j^{2(t-1)}\right)$$
4. Compute $n_j^{(t)} = \sum_{i=1}^n \mathbb{I}_{z_i^{(t)}=j}$, $s_j^{(t)} = \sum_{i=1}^n \mathbb{I}_{z_i^{(t)}=j} x_i$
5. Generate $p^{(t)}$ from $\mathrm{D}\left(\gamma_1 + n_1^{(t)}, \ldots, \gamma_k + n_k^{(t)}\right)$,
6. Generate $\mu_j^{(t)}$ $(j = 1, \ldots, k)$ from $\pi(\mu_j^{(t)}|z^{(t)}, x) = f_t(\mu_j|\mu_{*j}, s_{*j}, n_{*j}, \nu_{*j})$
7. Generate $\sigma_j^{2(t)}$ $(j = 1, \ldots, k)$ from $\pi(\sigma_j^{2(t)}|x, z^{(t)}) = f_{i\gamma}(\sigma_j^{2(t)}|\nu_{*j}, s_{*j})$.
8. goto 2 until $t = m$

2.2 Bayesian model selection: Chib's method for mixture of lognormal distributions

The estimation of the best approximating mixture model requires a choice of the optimal number of mixture components. The number of components is usually determined from the data, often by minimising the Bayesian Information Criteria (BIC). From a Bayesian perspective, the choice of the optimal number of mixture components is based on the maximisation of the marginal likelihood (ml) obtained by integrating the likelihood function with respect to the prior density (Gelfand and Dey 1994, Newton and Raftery 1994, Kass and Raftery 1995, Chib 1995). Newton and Raftery (1994) showed that the ml can be estimated as the harmonic mean of the likelihood values. This estimate is simulation consistent, but however not stable, because the inverse likelihood does not have a finite variance. Gelfand and Dey (1994) proposed a quantity for ml which is also consistent but requires a tuning function, which can be quite difficult to determine. To overcome this problem, Chib (1995) has developed a method for ml in the setting where the Gibbs sampling algorithm has been used to provide a sample of draws from the posterior distribution. We shall adopt his method for lognormal mixtures.

Let us first recall the expression of the BIC which was devised by Schwarz (1978) as an asymptotic approximation to the log integrated likelihood

$$BIC(k) = \log L(x|\hat{p}_j, \hat{\mu}_j, \hat{\sigma}_j^2) - \frac{\tau_k}{2}\log(n), \tag{8}$$

where τ_k is the number of free parameters of the model with k components and the $\hat{p}_j$, $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ are the classical estimates.

Chib's approach to compute the ml, $m(x|M_k)$ makes use of the Gibbs output

$$\left(p^{(g)}, \mu^{(g)}, \sigma^{2^{(g)}}, z^{(g)}\right)_{g=1}^{G},$$

from the set of the following complete conditional densities,

$$\pi(p|x,\mu,\sigma^2,z), \quad \pi(\mu,\sigma^2|x,p,z) \quad \text{and} \quad \pi(z|x,p,\mu,\sigma^2).$$

Since p and (μ, σ^2) are independent a priori, then given z, we have the following simplification

$$\pi(p|x,\mu,\sigma^2,z) = \pi(p|x,z) \qquad \pi(\mu,\sigma^2|x,p,z) = \pi(\mu,\sigma^2|x,z).$$

The posterior density, $\pi(p,\mu,\sigma^2|x)$ is a direct application of Bayes theorem,

$$\pi(p,\mu,\sigma^2|x) = \frac{L(x|p,\mu,\sigma^2)\pi(p,\mu,\sigma^2)}{\int L(x|p,\mu,\sigma^2)\pi(p,\mu,\sigma^2)},$$

$m(x)$, by virtue of being the normalising constant of the posterior density, can be written as

$$m(x) = \frac{L(x|p,\mu,\sigma^2)\pi(p,\mu,\sigma^2)}{\pi(p|x)\pi(\mu,\sigma^2|x)}. \tag{9}$$

This relation is valid for all the values of the parameters and in particular for a chosen arbitrary set of values (p^*,μ^*,σ^{2^*}). The proposed estimate of the marginal density, on the computationally convenient logarithm scale is

$$\log\hat{m}(x) = \log L(x|p^*,\mu^*,\sigma^{2*}) + \log\pi(p^*,\mu^*,\sigma^{2^*}) - \log\hat{\pi}(p^*|x) - \log\hat{\pi}(\mu^*,\sigma^{2*}|x). \tag{10}$$

The two first elements of the right hand side are straightforward to compute. The computation of the last two elements is at the core of Chib's method. The following deconvolutions

$$\pi(\mu,\sigma^2|x) = \int \pi(\mu,\sigma^2|x,z)p(z|x)dz \quad \text{and} \qquad \pi(p|x) = \int \pi(p|x,z)p(z|x)dz,$$

suggest how the Gibbs output can be used as Monte Carlo estimates of $\pi(\mu,\sigma^2|x)$ and $\pi(p|x)$ at μ^*, σ^{2*}, p^* based on the following Rao-Blackwell estimates

$$\begin{aligned} \hat{\pi}(\mu^*,\sigma^{2*}|x) &= \frac{1}{G}\sum_{g=1}^{G}\pi(\mu^*,\sigma^*|x,z^{(g)}), & (11) \\ \hat{\pi}(p^*|x) &= \frac{1}{G}\sum_{g=1}^{G}\pi(p^*|x,z^{(g)}). & (12) \end{aligned}$$

To compute the marginal density by Chib's method, it is necessary that all integrating constants of the full conditional distributions in the Gibbs sampler be known. By including all integrating constants on the prior densities and the likelihood given above the joint posterior density of μ_j and σ_j^2 conditional on x and z become

$$\begin{aligned} \pi(\mu_j,\sigma_j^2|x,z) \quad = \quad & \sigma_j^{-(n+\nu_0+3)}\sqrt{\frac{n_0}{(2\pi)^{n+1}}}\frac{(s_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \times \\ & \exp\left(-\frac{1}{2\sigma_j^2}\left(s_0+s^2(z)+n_j(\mu_j-\bar{x}_j(z))^2+n_0(\mu_j-\mu_0)^2\right)+n_j\bar{x}_j(z)\right), \end{aligned}$$

therefore,

$$\log\pi(\mu,\sigma^2|x,z) = \sum_{j=1}^{k}\log\pi(\mu_j,\sigma_j^2|x,z). \tag{13}$$

The posterior density of p conditional on x and z is

$$\pi(p|x,z) = f_{Dir}(\gamma_1 + n_1, ..., \gamma_k + n_k) = \frac{\Gamma\left(\sum_{j=1}^{k} \gamma_j + n_j\right)}{\prod_{j=1}^{k} \Gamma(\gamma_j + n_j)} \prod_{j=1}^{k} p_j^{\gamma_j + n_j - 1}. \tag{14}$$

We use the following algorithm to compute the marginal likelihood

1. Fix μ^*, σ^*, p^* as the element of the Gibbs output which maximises the likelihood function
2. Generate $z^{(g)}$ from the conditional density $\pi(z|x)$
3. Compute respectively $\pi(\mu^*, \sigma^*|x, z^{(g)})$ and $\pi(p^*|x, z^{(g)})$ in (11) and (12) and average them to obtain $\hat{\pi}(\mu^*, \sigma^*|x)$, $\hat{\pi}(p^*|x)$.
4. Finally using the logarithm of the joint prior densities, the log likelihood of the mixture, compute the marginal likelihood given in (10)

2.3 Application to the income distribution in the UK

The data are from the Family expenditure Survey (FES) a continuous survey of samples of the UK population living in households. The data, which were also used by Flachaire and Nunez (2002) cover four waves of survey 1979, 1988, 1992 and 1996. They correspond to household disposable income (i.e post-tax and transfer income) before housing costs and are adjusted by the McClements adult-equivalence scale. We deflated the data by the corresponding relative consumer price index.

2.3.1 Sample characteristics and Prior selection

It is not possible to estimate the parameters of a mixture of densities without any prior information. Even in a classical framework, any software asks hints about the location and shape of the different components. In a Bayesian framework, there is a fundamental identification problem as in the Gibbs exploration the partition Z_j may contain empty cells which lead to an infinite likelihood. So the n_0, s_0, μ_0 cannot be zero. The usual practice consists in assigning the same prior to the k elements of the mixture. This information is usually based on sample moments (see for instance Marin and Robert 2007 using the mean and the variance). However, if this type of information solves the empty cell problem, it does not provide meaningful information on member location and shape and thus cannot avoid the phenomenon of label switching. The elicitation of a more precise prior information is needed. For each member, we have to elicit four parameters

- $\pi(\sigma^2)$: $E(\sigma^2) = s_0/(\nu_0 - 2)$
- $\pi(\mu|\sigma^2)$: $E(\mu) = \mu_0$, $Var(\mu|\sigma^2) = s_0/(\nu_0 - 2)/n_0$

It is possible to fix ν_0 and n_0 at predefined values such that $\nu_0 = 3$ and $n_0 = 1$. We have to rely on sample information to determine the remaining two parameters.

Let us first detail some properties of the log-normal density $\Lambda(x|\mu, \sigma^2)$. In this notation, μ happens to be equal to the log of the median, while $\sigma^2 = \log(1 + V_x/m_x^2)$ where V_x and m_x correspond respectively to the sample variance and to the sample mean. Noting $q_{0.50}$ for the median, we can determine a prior information for the central member of the mixture:

- $\mu_{0c} = \log(q_{0.50})$
- $s_{0c}/(\nu_0 - 2) = \log(1 + V_x/m_x^2)$

To elicit the parameters of the other members of the mixture, we estimate the quantiles of the distribution. For instance for a four component mixture, we estimate $q_{0.25}$, $q_{0.50}$, $q_{0.75}$ and $q_{0.95}$. The μ_{0i} will be given by $\mu_{0i} = \log(q_i)$ so that there are ordered, while the s_{0i} are kept constant and equal to s_{0c}. For the p_i, we keep a uniform prior. Compared to the usual method, only the mu_{0i} are different as they are ordered according to the selected quantiles.

We are going to try to fit two, three and four component mixtures. We get the following prior information displayed in Table 3 for the four samples.

Table 1: Prior information

	μ_{01}	μ_{02}	μ_{03}	μ_{04}	s_{01}	s_{02}	s_{03}	s_{04}
1979	3.96	4.31	4.64	5.06	0.22	0.22	0.22	0.22
1988	4.05	4.47	4.85	5.37	0.35	0.35	0.35	0.35
1992	4.09	4.52	4.92	5.46	0.39	0.39	0.39	0.39
1996	4.19	4.55	4.92	5.44	0.32	0.32	0.32	0.32

In the usual case depicted in Marin and Robert (2009), the Chib's method has to be modified in order to take into account the possibility of label switching. With an informative prior, this modification is no longer needed.

2.3.2 Posterior results

We use the algorithm described above to generate 1,000 draws from the posterior density in order to select the optimal number of components by computing the BIC and the marginal likelihood (ml) with the Chib's method for $k = 2$, $k = 3$ and $k = 4$. After, using the selected model we estimate the parameters of each group using 10,000 draws.

Table 2 shows that the BIC and the marginal likelihood are respectively minimised and maximised when $k = 3$ for the four years of FES data; the best approximating model is a mixture of 3 lognormal distributions.

Table 2: Choice of the optimal number of component of the mixture model based on 1000 draws

k		2	3	4
1979	ml	-20927.81	**-18660.38**	-19961.83
	BIC	61549.46	**61386.89**	61406.67
1988	ml	-25068.22	**-23143.29**	-24226.88
	BIC	68106.60	**67922.99**	67951.60
1992	ml	-27840.53	**-25010.59**	-25617.08
	BIC	71053.82	**70839.11**	70866.50
1996	ml	-23268.17	**-22006.88**	-22936.87
	BIC	64554.93	**64436.08**	64453.25

Table 3 gives the estimates of posterior mean of parameters associated with their corresponding standard deviations. Combining these results with the estimated group means and standard deviations given in Table 4, we can identify that the first member of the mixture corresponds to poor people, those with the lowest income. They correspond to 23% of the sample on average. The second member of the mixture can be identified to the middle class, with middle mean incomes, representing on average 73% of the sample. The last mixture member has the highest mean and thus can be identified to higher incomes, representing on average 4% of the sample. Inequality within and between these groups will be analysed in section 4.

3 Bayesian inference for inequality from Gibbs output

In this section, we detail how an inequality index can be implemented in the framework of a mixture of distributions. We shall see that decomposability is a highly simplifying property. The class of Generalised Entropy indices have this property while the Gini index, for instance, has not.

Table 3: **Posterior inference for FES data**

based on 10 000 draws

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$
1979	3.83	4.46	4.12	0.030	0.15	1.07	0.222	0.756	0.022
	(0.016)	(0.022)	(0.237)	(0.004)	(0.013)	(0.281)	(0.030)	(0.043)	(0.015)
1988	3.94	4.63	4.50	0.038	0.21	1.20	0.219	0.737	0.044
	(0.010)	(0.017)	(0.094)	(0.004)	(0.011)	(0.210)	(0.018)	(0.025)	(0.013)
1992	3.99	4.74	4.44	0.058	0.20	1.26	0.264	0.676	0.059
	(0.022)	(0.034)	(0.089)	(0.008)	(0.019)	(0.192)	(0.037)	(0.04)	(0.012)
1996	4.12	4.72	4.37	0.070	0.22	2.21	0.249	0.734	0.017
	(0.024)	(0.036)	(0.210)	(0.010)	(0.014)	(0.564)	(0.049)	(0.052)	(0.005)

Table 4: Estimated mean and variance for FES data

	Mean			S.D.		
	Group1	Group2	Group3	Group1	Group2	Group3
1979	46.76	93.22	105.11	8.15	37.50	145.46
1988	52.41	113.86	164.02	10.31	55.04	249.83
1992	55.65	126.47	159.17	13.59	59.50	252.95
1996	63.75	125.21	238.65	17.16	62.11	679.86

3.1 Decomposability of inequality indices

Decomposability is a very useful properties for indices, both for interpretation and inference when within mixtures. A decomposable inequality index can be expressed as a weighted average of inequality within subgroups, plus inequality among those subgroups.

Let $I(x, n)$ be the inequality value for a population of n individuals with income distribution x. $I(x, n)$ is assumed to be continuous and symmetric in x, $I(x, n) \geq 0$ with perfect equality holding if and only if $x_i = \mu$ for all i, and $I(x, n)$ is supposed to have a continuous first order partial derivative.

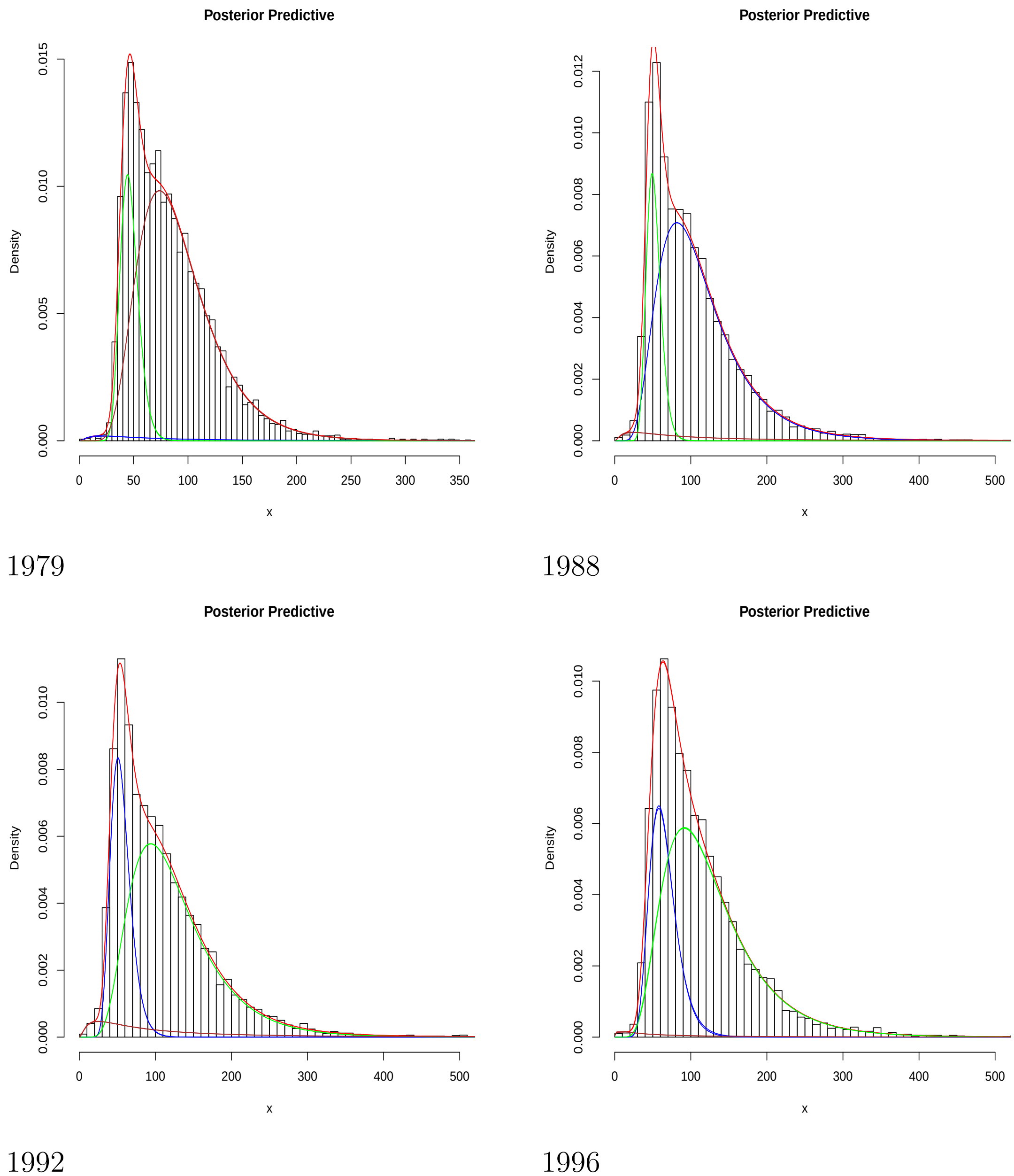
Under these assumptions, additive decomposition condition can be stated as follows (Shorrocks 1983):

Definition 1. *Given a population of any size $n \geq 2$ and a partition into k non-empty subgroups, the inequality index $I(x, n)$ is decomposable if there exists a set coefficients $\tau_j^k(\mu, n)$ such that*

$$I(x, n) = \sum_{j=1}^{k} \tau_j^k I(x^j; n_j) + B,$$

where $x = (x^1, \ldots, x^k)$, $\mu = (\mu_1, \ldots, \mu_k)$ is the vector of subgroup means $\tau_j(\mu, n)$ is the weight attached to subgroup j in a decomposition into k subgroups, and B is the

Figure 1: Density inference for the FES data



between-group term, assumed to be independent of inequality within the individual *subgroups.*

Making within-group transfers until x_i^j in each subgroup and letting u_n represent the unit vector with n components, we obtain $B = I(\mu_1 u_{n_1}, \ldots, \mu_k u_{n_k})$.

The GE indices (see Cowell 1977 or Bourguignon 1979) are given by

$$I_{GE}^{\alpha} = \frac{1}{\alpha^2 - \alpha} \int \left[\left(\frac{x}{\mu(F)} \right)^{\alpha} - 1 \right] f(x) dx \tag{15}$$

where $\alpha \in (-\infty, +\infty)$ is a parameter that captures the sensitivity of a specific GE index to particular parts of the distribution: for α large and positive the index is sensitive to changes in the distribution that affect the upper tail; for α negative the index is sensitive to changes in distribution that affect the lower tail. In empirical works, the range of values for α is typically restricted to $[-1, 2]$ (see Shorrocks 1980) because, otherwise, estimates may be unduly influenced by a small number of very small incomes or very high incomes.

For the lognormal distribution, the GE indices provide an analytical expression which is given by

$$I_{GE}^{\alpha} = \frac{\exp[(\alpha^2 - \alpha)\sigma^2/2] - 1}{\alpha^2 - \alpha}.$$

The GE family for mixture model with k components in $f(.)$ is given

$$\begin{aligned}
I_{GE}^{\alpha} &= \frac{1}{\alpha^2 - \alpha} \int \left[\left(\frac{x}{\sum_{j=1}^k p_j \mu_j} \right)^{\alpha} - 1 \right] \sum_{j=1}^k p_j f_j(x) dx \\
&= \sum_{j=1}^k p_j \frac{1}{\alpha^2 - \alpha} \int \left[\left(\frac{x \mu_j}{\mu_j \sum_{j=1}^k p_j \mu_j} \right)^{\alpha} - 1 \right] f_j(x) dx, \\
&= \sum_{j=1}^k p_j \left(\frac{\mu_j}{\sum_{j=1}^k p_j \mu_j} \right)^{\alpha} \frac{1}{\alpha^2 - \alpha} \int \left[\left(\frac{x}{\mu_j} \right)^{\alpha} - 1 \right] f_j(x) dx \\
&\quad + \frac{1}{\alpha^2 - \alpha} \left[\sum_{j=1}^k p_j \left(\frac{\mu_j}{\sum_{j=1}^k p_j \mu_j} \right)^{\alpha} - 1 \right],
\end{aligned}$$

If $\tau_j = p_j \mu_j / \sum_{j=1}^k p_j \mu_j$ and I_{GE}^j denotes the generalised entropy family index with parameter α for the group j then

$$I_{GE}^{\alpha} = \underbrace{\sum_{j=1}^k p_j^{1-\alpha} \tau_j^{\alpha} I_{GE}^j}_{withinGE} - \underbrace{\frac{1}{\alpha^2 - \alpha} \left(\sum_{j=1}^k p_j^{1-\alpha} \tau_j^{\alpha} - 1 \right)}_{betweenGE}. \tag{16}$$

Measures ordinally equivalent to the GE class include a number of pragmatic indexes such as the mean logarithmic deviation index (I_{MLD}) ($I_{MLD} = \lim_{\alpha \to 0} I_{GE}^{\alpha}$), Theil's index ($I_{Theil} = \lim_{\alpha \to 1} I_{GE}^{\alpha}$) and the coefficient of variation ($1/2 I_{CV}^2 = \lim_{\alpha \to 2} I_{GE}^{\alpha}$). The most popular variants of this specific class of GE family are the Theil and the MLD since they are the only zero homogeneous decomposable measures such that the weights of the within-group-inequalities in the total inequality sum to a constant (see Bourguignon 1979).

$$
\begin{aligned}
I_{Theil}(F) &= \int \frac{x}{\mu(F)} log\left(\frac{x}{\mu(F)}\right) f(x)dx, \\
I_{MLD}(F) &= -\int log\left(\frac{x}{\mu(F)}\right) f(x)dx,
\end{aligned}
$$

and are expressed for a mixture point of view as

$$
I_{Theil} = \underbrace{\sum_{j=1}^{k} \tau_j I_{Theil}^{j}}_{withinTheil} + \underbrace{\sum_{j=1}^{k} \tau_j log\left(\frac{\tau_j}{p_j}\right)}_{betweenTheil}. \tag{17}
$$

$$
I_{MLD} = \underbrace{\sum_{j=1}^{k} p_j I_{MLD}^{j}}_{withinMLD} - \underbrace{\sum_{j=1}^{k} p_j log\left(\frac{\tau_j}{p_j}\right)}_{betweenMLD}. \tag{18}
$$

For the mixture of lognormal distributions, we have the equality

$$
I_{Theil}^{j} = I_{MLD}^{j} = \frac{\sigma_j^2}{2} \qquad \text{and} \qquad \tau_j = \frac{p_j \exp(\mu_j + \sigma_j^2/2)}{\sum p_j \exp(\mu_j + \sigma_j^2/2)}
$$

The Atkinson index is expressed as

$$
I_A^{\epsilon} = 1 - \frac{1}{\mu(F)} \left[\int x^{1-\epsilon} dF(x)\right]^{\frac{1}{1-\epsilon}}, \tag{19}
$$

where $\epsilon \geq 0$ is a parameter defining (relative) inequality aversion. A brief comparison of Atkinson's and the GE's measure shows that they have different cardinalisation functions but they are ordinally equivalent for cases $\alpha \leq 1$ and $\epsilon = 1 - \alpha$ so that we have the relation:

$$
I_A^{\epsilon}(F) = 1 - \frac{1}{\mu(F)} \left[(\alpha^2 - \alpha) I_{GE}^{\alpha}(F) + 1\right]^{\frac{1}{\alpha}}. \tag{20}
$$

Consequently, the Atkinson index is decomposable, thanks to the properties of the GE, but this decomposition is an indirect one (see Cowell 1977).

3.2 Bayesian inference for inequality indices

The Generalised Entropy and the Atkinson indices are simple transformations of the parameters in a mixture of lognormals. We thus simulate the posterior density of these indices directly from the Gibbs output $\left(p^{(g)}, \mu^{(g)}, \sigma^{2^{(g)}}\right)_{g=1}^{G}$.

Despite the fact that the Gini index is not decomposable and thus not easily computed for a mixture of distributions, it is still interesting to compute it numerically, just for the sake of comparison. We rely on the numerical approximation:

$$\hat{I}_G^{(g)} = \frac{1}{\mu} \int_0^{\infty} \Phi(x|p^{(g)}, \mu^{(g)}, \sigma^{2^{(g)}})(1 - \Phi(x|p^{(g)}, \mu^{(g)}, \sigma^{2^{(g)}}))dx,$$

where $\Phi(.)$ is the cumulative distribution function of the mixture of lognormal distributions.

With either method, we get G draws for each inequality index, we take the mean and the standard deviation to obtain reliable estimation for each of them.

4 Inequality growth in UK from 1979 to 1996

Since the late 1970s, the gap between the rich and the poor in the UK has considerably increased. It is now recognised that the growth of inequality in Great Britain has been faster than in any comparable industrial countries (see for instance Jenkins 2000).

4.1 Total inequality

In Table 5, we have estimated various inequality indices over the four sample periods: the Generalised Entropy index for $\alpha = 0.5$, Theil, Mean Logarithmic Deviation (MLD), the Atkinson index with $\epsilon = 0.5$ and Gini indices, with their standard errors in brackets. There is a considerable increase of all these inequality indices from 1979 to 1988, to be put in relation to the period (1979-1990) when Margaret Thatcher was Prime Minister. This inequality growth is slowing down between 1988 and 1992. And from 1992 to 1996, all these inequality measures decrease as confirmed by Figure 2. Using the same data, the annual report of the Department of Social Security (1998) have found comparable results. By decomposing inequality, we shall see that this general picture is not so simple.

Table 5: Estimates and Standard errors of inequality indices

	GE ($\alpha = 0.5$)	Theil	MLD	Atkinson ($\alpha = 0.5$)	Gini
1979	0.108	0.104	0.107	0.054	0.255
	(0.0041)	(0.0027)	(0.0027)	(0.0022)	(0.0021)
1988	0.168	0.148	0.160	0.0832	0.307
	(0.0093)	(0.0043)	(0.0057)	(0.0048)	(0.0032)
1992	0.182	0.165	0.176	0.0898	0.321
	(0.0086)	(0.0042)	(0.0057)	(0.0044)	(0.0035)
1996	0.168	0.150	0.146	0.0739	0.295
	(0.0087)	(0.0097)	(0.0095)	(0.0090)	(0.0033)

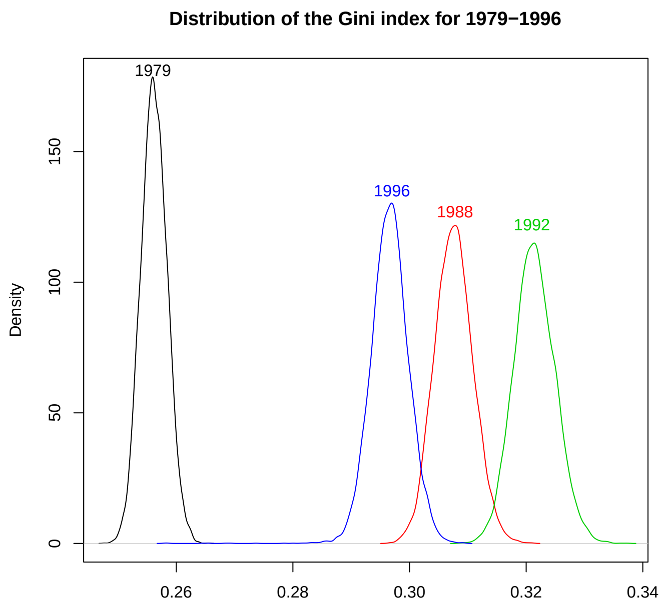


Figure 2: Evolution of the Gini index

4.2 Inequality decomposition

We have identified three groups of income, each one corresponding to one member of the estimated mixture. We can decompose inequality using this structure. The weights of the mixture give information on the evolution of the importance of each group (see Table 3). The importance of the low income group remained constant between 1979 and 1988 (0.22), slightly increased in 1992 (0.26) and slightly decreased in 1996 (0.25). The importance of the middle income group decreased continuously from 1979 (0.76) to 1992 (0.68) and regained some importance in 1996 (0.73), roughly back to its level of 1988. The importance of the upper income group experiences the largest variation. Its weight was initially very small (0.022). It doubles in 1988 (0.044), continue to increase in 1992 (0.059), but experienced a tremendous drop in 1996 (0.017), reaching a weight smaller than that of 1979. There thus has been a large change in structure of the three groups. The importance of the low income group did not change much, that of the high income was the most affected by the fight against inequality that followed the Thatcher period.

After examining the structural evolution of incomes groups, we compute the decomposition of within versus between inequality. Within inequality represents on average 68% of total inequality and this proportion does not vary too much. Within group inequality increased from 1979 to 1988 and then remained roughly constant. Between group inequality increased from 1979 till 1992 and decreases only after this date.

Table 6: MLD decomposition of income inequality

Year	Within	Prop.	Between	Prop.	Total
1979	0.072	0.67	0.036	0.33	0.108
	(0.0036)		(0.0034)		(0.0028)
1988	0.111	0.70	0.048	0.30	0.159
	(0.0048)		(0.0048)		(0.0054)
1992	0.115	0.66	0.059	0.34	0.175
	(0.0059)		(0.0055)		(0.0052)
1996	0.108	0.70	0.047	0.30	0.155
	(0.0075)		(0.0100)		(0.0101)

Finally, with Table 7, we can trace the evolution of within inequality, for each of the three income groups.

Within the poor group, inequality was roughly multiplies by 10 between 1979 and 1988. It then decreased steadily to recover in 1996 roughly the same level as in 1979. *Within the middle income group*, inequality remains constant between 1979 and 1988. It strongly increased after that date, especially between 1992 and 1996. Finally, inequality *within the upper income group* decreased between 1979 and 1988, then increased and again decreased.

Figure 3: MLD decomposition of income inequality changes in UK over the years 1979, 1988, 1992, 1996

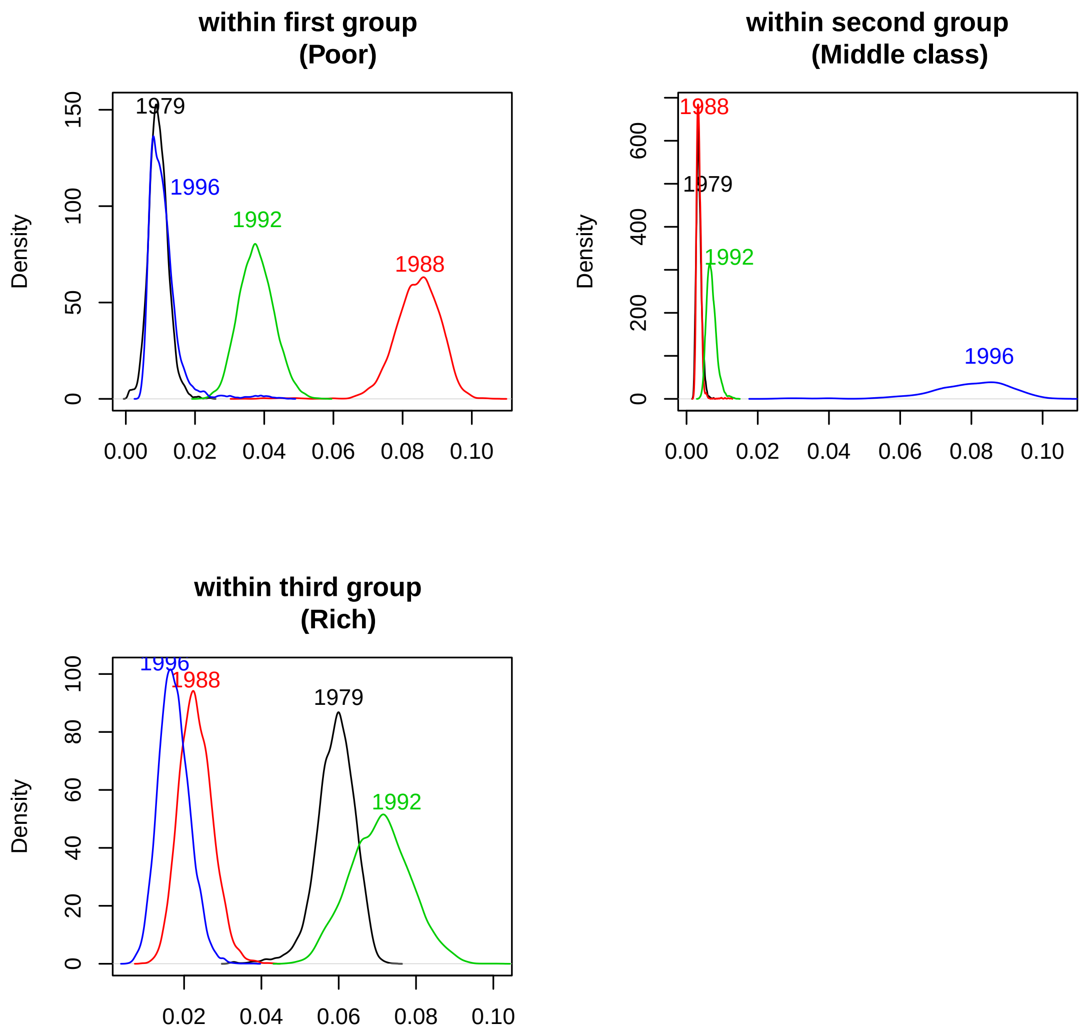


Table 7: Generalised Entropy Family decomposition of income inequality changes in the UK

	MLD Within-group Inequality		
Year	Poor	Medium	High
1979	0.0094	0.0035	0.0591
	(0.0028)	(0.0007)	(0.0050)
1988	0.0843	0.0035	0.0231
	(0.0067)	(0.0035)	(0.0231)
1992	0.0378	0.0070	0.0706
	(0.0050)	(0.0014)	(0.0078))
1996	0.0109	0.0793	0.0174
	(0.0049)	(0.0122)	(0.0040)

5 Conclusion

The mixture of lognormal distributions chosen is found to be a convenient explanatory model of income distribution. Tested on the UK FES, it clearly manages to identify income groups. The class of Generalised Entropy inequality indices can fit in this mixture framework, thanks to its additive decomposability that follows nicely the additive structure of mixtures.

We have demonstrated how a Gibbs sampler can be used to estimate the mixture of lognormal distributions when we elicit a more precise prior information to avoid the usual "label switching" problem. We were able to give a general rule of elicitation.

The Bayesian approach for estimating mixtures has allowed us to shed some light on the small sample properties of inequality measures.

References

Aitchison, J. and Brown, J. A. C. (1957). *The lognormal Distribution.* Cambridge University Press, London.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, 47:901–920.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1321.

Chotikapanich, D. and Griffiths, W. E. (2008). Estimating income distributions using a mixture of gamma densities. In Chotikapanich, D., editor, *Modeling Income distributions and Lorenz Curves*, pages 285–302. Springer, Berlin.

Cowell, F. A. (1977). *Measuring Inequality.* Philip Allan Publishers Limited, Oxford.

Flachaire, E. and Nunez, O. (2002). Estimation of the income distribution and detection of subpopulations: An explanatory model. *Computational Statistics and Data Analysis*, 51:3368–3380.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society*, 56:501–514.

Ghosal, S. and Van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities. *Annals of Statistics*, 29(5):1233–1263.

Jenkins, S. P. (1996). Recent trends in the uk income distribution: What happened and why? *Oxford Review of Economic Policy*, 12(1):29–46.

Jenkins, S. P. (2000). Trends in the uk income distribution. In Hauser, R. and Becker, I., editors, *The personal Distribution of Income in an International Perspective.* Springer-Verlag, Berlin.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association*, 90:773–795.

Marin, J. and Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics.* Springer Texts in Statistics, XIII, 255 pp.

Marin, J.-M. and Robert, C. P. (2009). Importance sampling methods for Bayesian discrimination between embedded models. Technical report, Université Paris Dauphine.

Marron, J. S. and Schmitz, H. P. (1992). Simultaneous density estimation of several income distributions. *Econometric Theory*, 8:476–488.

McDonald, J. and Ransom, M. (1979). Functional forms, estimation techniques and the distribution of income. *Econometrica*, 47:1513–1525.

McDonald, J. B. (1984). Some generalized functions for the size distribution income. *Econometrica*, 52:647–663.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society*, 56:3–48.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3):613–625.

Singh, S. K. and Maddala, G. S. (1976). A functions for size distribution of incomes. *Econometrica*, 44(5):963–970.